A RECURSIVE NONPARAMETRIC ESTIMATOR FOR THE TRANSITION KERNEL OF A PIECEWISE-DETERMINISTIC MARKOV PROCESS

ROMAIN AZAÏS

Abstract. In this paper, we investigate a nonparametric approach to provide a recursive estimator of the transition density of a non-stationary piecewise-deterministic Markov process, from only one observation of the path within a long time. In this framework, we do not observe a Markov chain with transition kernel of interest. Fortunately, one may write the transition density of interest as the ratio of the invariant distributions of two embedded chains of the process. Our method consists in estimating these invariant measures. We state a result of consistency under some general assumptions about the main features of the process. A simulation study illustrates the well asymptotic behavior of our estimator.

1. Introduction

The purpose of this paper is to investigate a nonparametric recursive method for estimating the transition kernel of a non-stationary piecewise-deterministic Markov process, from only one observation of the process within a long time interval.

Piecewise-deterministic Markov processes (PDMP's) have been introduced in the literature by Davis in [7]. They are a general class of non-diffusion stochastic models involving deterministic motion broken up by random jumps, which occur either when the flow reaches the boundary of the state space or in a Poisson-like fashion. The path depends on three local features namely the flow Φ, which controls the deterministic trajectories, the jump rate λ, which governs the inter-jumping times, and the transition kernel Q, which determines the post-jump locations. An appropriate choice of the state space and the main characteristics of the process covers a large variety of stochastic models covering problems in reliability (see [7] and [5]) or in biology (see [13] and [10]) for instance. In this context, it appears natural to propose some nonparametric methods for estimating both the characteristics λ and Q, which control the randomness of the motion. Indeed, the deterministic flow is given by physical equations or deterministic biological models. In [2], Azaïs *et al.* proposed a kernel method for estimating the conditional probability density function associated with the jump rate λ, for a non-stationary PDMP defined on a general metric space. This work was based on a generalization of Aalen's multiplicative intensity model and a discretization of the state space. In the present paper, we assume that Q admits a density q with respect to the Lebesgue measure, and we focus on the nonparametric estimation

2010 *Mathematics Subject Classification.* Primary: 62G05, Secondary: 62M05.

Key words and phrases. Piecewise-deterministic Markov processes, nonparametric estimation, recursive estimator, transition kernel, asymptotic consistency.

This work was supported by ARPEGE program of the French National Agency of Research (ANR), project "FAUTOCOES", number ANR-09-SEGI-004.

of this function, from the observation of a PDMP within a long time, without assumption of stationarity. In addition, since measured data are often processed sequentially, it is convenient to propose a recursive estimator.

Nonparametric estimation methods for stationary Markov chains have been extensively investigated, beginning with Roussas in [20]. He studied kernel methods for estimating the stationary density and the transition kernel of a Markov chain satisfying the strong Doeblin's condition. Later, Rosenblatt proposed in [19] some results on the bias and the variance of this estimator in a weaker framework. Next, Yakowitz improved in [22] the previous asymptotic normality result assuming a Harris-condition. Masry and Györfi in [17], and Basu and Sahoo in [3], have completed this survey. There exists also an extensive literature on nonparametric estimates for non-stationary Markov processes. We do not attempt to present an exhaustive survey on this topic, but refer the interested reader to [6, 8, 9, 11, 14, 15, 16] and the references therein. In this new framework, Doukhan and Ghindès have investigated in [8] a bound of the integrated risks for their estimate. Hernández-Lerma *et al.* in [11] and Duflo in [9] made inquiries about recursive methods for estimating the transition kernel or the invariant distribution of a Markov chain. Liebscher gave in [16] some results under a weaker condition than Doeblin's assumption. More recently, Clémençon in [6] proposed a quotient estimator using wavelets and provided the lower bound of the minimax $\mathbf{L}^p$-risk. Lacour suggested in [15] an estimator by projection with model selection, next she introduced in [14] an original estimate by inquiring into a new contrast derived from regression framework.

Our investigation and the studies of the literature mentioned before are different and complementary. In this paper, we propose to estimate the transition density q of a PDMP by kernel methods. Nevertheless, we do not observe a Markov chain whose transition distribution is given by q. Fortunately, one may write the function of interest as the ratio of two invariant measures: the one of the two components pre-jump location and post-jump location, over the one of the pre-jump location. Indeed, $Q(x, A)$ is defined as the conditional probability that the post-jump location is in A, given the path is in x just before the jump. Therefore, we suggest to estimate both these invariant measures in order to provide an estimator of the transition kernel Q. A major stumbling block for estimating the invariant law of the pre-jump location is related with the transition kernel of this Markov chain, which may charge the boundary of the state space. As a consequence, the transition kernel, as well as the corresponding invariant distribution, admits a density only on the interior of the state space. The investigated approach for estimating the invariant measure is based on this property of the transition kernel. But the main difficulty appears for analyzing the two-components process pre-jump location, post-jump location. This Markov chain has a special structure, because its invariant distribution admits a density function on the interior of the state space, unlike its transition kernel. Indeed, the pre-jump location is distributed on the curve governed by the deterministic flow initialized by the previous post-jump location. As a consequence, the author have to explore a new method for estimating the two-dimensional invariant measure of interest. The proposed one is more universal, but implies a more restrictive assumption on the shape of the bandwidth.

An intrinsic complication throughout the paper comes from the presence of deterministic jumps, when the path tries to cross the boundary of the state space. Indeed, this induces that the invariant distributions mentioned above may charge a subset with null Lebesgue measure. This important feature has been introduced by Davis in [7] and is very attractive

for the modeling of a large number of applications. For instance, one may find in [1] an example of shock models with failures of threshold type. One may also refer the reader to [13], where the authors develop a PDMP to capture the mechanism behind antibiotic released by a bacteria. Forced jumps are used to model a deterministic switching when the concentration of nutrients rises over a certain threshold.

The paper is organized as follows. Section 2 is devoted to the precise formulation of our problem. The sequel of the article is divided into three parts. Section 3 is devoted to a recursive kernel estimator of the invariant measure of the pre-jump location. More precisely, in Subsection 3.1, we examine the ergodicity of the Markov chain of the pre-jump locations. Next, in Subsection 3.2, we introduce a nonparametric estimator of the corresponding invariant distribution and we prove its almost sure convergence in Proposition 3.11. In Section 4, we investigate a nonparametric recursive method for estimating the invariant distribution of the two-components chain pre-jump location, post-jump location. First, some properties of this Markov process are studied in Subsection 4.1. The recursive estimator is provided in Subsection 4.2 and the associated convergence is stated in Proposition 4.7. Our nonparametric estimator of the function of interest q is defined as the ratio of both the previously mentioned recursive estimators. The main result of consistency lies in Theorem 4.8. Finally, Section 5 deals with numerical considerations for illustrating the asymptotic behavior of our estimate.

2. Preliminaries

This section is devoted to the definition of a piecewise-deterministic Markov process on $\mathbf{R}^d$, where d is an integer greater or equal to 1. The process evolves in an open subset E of $\mathbf{R}^d$ equipped with the Euclidean norm $|\cdot|$. The motion is defined by the three local characteristics (λ, Q, Φ).

- $\Phi : \mathbf{R}^d \times \mathbf{R} \to \mathbf{R}^d$ is the deterministic flow. It satisfies,
$$\forall \xi \in \mathbf{R}^d,\ \forall s, t \in \mathbf{R},\ \Phi_\xi(t+s) = \Phi_{\Phi_\xi(t)}(s).$$
For each $\xi \in E$, $t^+(\xi)$ denotes the deterministic exit time from E:
$$t^+(\xi) = \inf\{t > 0 \ :\ \Phi_\xi(t) \in \partial E\},$$
with the usual convention $\inf \emptyset = +\infty$.
- $\lambda : \mathbf{R}^d \to \mathbf{R}_+$ is the jump rate. It is a measurable function which satisfies,
$$\forall \xi \in \mathbf{R}^d,\ \exists \varepsilon > 0,\ \int_0^\varepsilon \lambda\big(\Phi_\xi(s)\big)\mathrm{d}s < +\infty.$$
- Q is a Markov kernel on $(\mathbf{R}^d, \mathcal{B}(\mathbf{R}^d))$ which satisfies,
$$\forall\ \xi \in \overline{E},\ Q(\xi, \overline{E} \setminus \{\xi\}) = 1 \quad \text{and} \quad Q(\xi, E) = 1.$$

There exists a filtered probability space $(\Omega, \mathcal{A}, (\mathcal{F}_t), \mathbf{P})$, on which the process (X_t) is defined (see [7]). Starting from $x \in E$, the motion can be described as follows. T_1 is a positive random variable whose survival function is,
$$\forall t \geq 0,\ \mathbf{P}(T_1 > t | X_0 = x) = \exp\left(-\int_0^t \lambda(\Phi_x(s))\mathrm{d}s\right) \mathbf{1}_{\{0 \leq t < t^+(x)\}}.$$

One chooses an E-valued random variable Z_1 according to the distribution $Q(\Phi_x(T_1), \cdot)$. Let us remark that the post-jump location depends only on the pre-jump location $\Phi_x(T_1)$. The trajectory between the times 0 and T_1 is given by

$$X_t = \begin{cases} \Phi_x(t) & \text{for } 0 \leq t < T_1, \\ Z_1 & \text{for } t = T_1. \end{cases}$$

Now, starting from X_{T_1}, one selects the time $S_2 = T_2 - T_1$ and the post-jump location Z_2 in a similar way as before, and so on. This gives a strong Markov process with the T_k's as the jump times (with $T_0 = 0$). One often considers the embedded Markov chain (Z_n, S_n) associated to the process (X_t) with $Z_n = X_{T_n}$, $S_n = T_n - T_{n-1}$ and $S_0 = 0$. The Z_n's denote the post-jump locations of the process, and the S_n's denote the interarrival times.

In this paper, we assume that the transition kernel Q admits a density q. Our main objective is the estimation of this function from the observation of one trajectory of the process within a long time.

Assumption 2.1. *We assume that the transition kernel Q admits a density q. That is,*

$$\forall \xi \in \mathbf{R}^d,\ \forall B \in \mathcal{B}(\mathbf{R}^d),\ Q(\xi, B) = \int_B q(\xi, z)\mathrm{d}z.$$

In addition, q is assumed to be piecewise-continuous.

In the following, we shall consider the discrete-time process (Z_n^-) defined by,

$$\forall n \geq 1,\ Z_n^- = \Phi_{Z_{n-1}}(S_n).$$

This sequence is naturally of interest. Indeed, the transition kernel Q describes the transition from Z_n^- to Z_n. Z_n^- stands for the location of (X_t) just before the nth jump. We shall prove that (Z_n^-) is a Markov chain in Lemma 3.1.

Our main objective in this paper is to provide a recursive estimator of $q(x, y)$. The recursive estimator of $q(x, y)$ that we consider may be written as follows,

$$\widehat{q}_n(x, y) = \frac{\displaystyle\sum_{j=1}^{n+1} \frac{1}{w_j^{2d}} K\left(\frac{Z_j^- - x}{w_j}\right) K\left(\frac{Z_j - y}{w_j}\right)}{\displaystyle\sum_{j=1}^{n+1} \frac{1}{v_j^d} K\left(\frac{Z_j^- - x}{v_j}\right)},$$

where $w_j = w_1 j^{-\beta}$, $v_j = v_1 j^{-\alpha}$, with $\alpha, \beta > 0$, and K is a kernel function satisfying Assumptions 3.9. Under some assumptions imposed in the sequel, we will state our major result of consistency in Theorem 4.8, that is

$$\widehat{q}_n(x, y) \xrightarrow{a.s.} q(x, y), \quad \text{as } n \text{ goes to infinity.}$$

Let us introduce some notations. In all the sequel, f and G denote the conditional probability density function and the conditional survival function associated with $\lambda(\Phi_\cdot(\cdot))$. Precisely, for all $z \in \mathbf{R}^d$ and $t \geq 0$,

$$\begin{aligned} G(z, t) &= \exp\left(-\int_0^t \lambda(\Phi_z(s))\mathrm{d}s\right), \\ f(z, t) &= \lambda(\Phi_z(t))G(z, t). \end{aligned}$$

In addition, throughout the paper, $\mathcal{S}$ denotes the conditional distribution of S_{n+1} given Z_n, for all integer n. For all $z \in E$ and $\Gamma \in \mathcal{B}(\mathbf{R}_+)$, we have

$$\begin{aligned} \mathcal{S}(z,\Gamma) &= \mathbf{P}\left(S_{n+1}\in\Gamma | Z_n = z, \sigma(Z_i, S_i \,:\, 0\leq i\leq n)\right) \\ &= \int_{\Gamma\cap[0,t^+(z)[} f(z,s)\mathrm{d}s \;+\; \mathbf{1}_\Gamma(t^+(z))G(z,t^+(z)). \end{aligned} \tag{1}$$

The relation between $\mathcal{S}$ and the conditional survival function G is given, for all $z \in \mathbf{R}^d$ and $t \geq 0$, by $G(z,t) = \mathcal{S}(z,]t,+\infty[)$.

In the next section, we focus on the invariant distribution of the process (Z_n^-) and we investigate a nonparametric way to estimate it.

3. Estimation of the invariant distribution of (Z_n^-)

The main objective of this section is the estimation of the invariant distribution of the Markov chain (Z_n^-). This section is divided into two parts. In the first one, we are interested in the existence and the uniqueness of the invariant distribution of (Z_n^-), and in the properties of its transition kernel $\mathcal{R}$. In the second part, we propose a recursive estimator of the invariant distribution of (Z_n^-) and we investigate its asymptotic behavior.

3.1. Some properties of (Z_n^-).

In this part, we focus on the process (Z_n^-), which is a Markov chain on $\overline{E}$. We especially investigate its transition kernel $\mathcal{R}$ and the existence of an invariant measure.

Lemma 3.1. *(Z_n^-) is a Markov chain whose transition kernel $\mathcal{R}$ is given, for all $y \in E$ and $B \in \mathcal{B}(\overline{E})$, by*

$$\mathcal{R}(y,B) = \int_E Q(y,\mathrm{d}z)\mathcal{S}(z,\Phi_z^{-1}(B)\cap\mathbf{R}_+), \tag{2}$$

where the conditional distribution $\mathcal{S}$ has already been defined by (1).

Proof. For all integer n, by (1), we have

$$\begin{aligned} &\mathbf{P}(Z_{n+1}^- \in B | Z_n = z, Z_n^-, \ldots, Z_1^-) \\ &\qquad = \mathbf{P}(S_{n+1}\in\Phi_z^{-1}(B)\cap\mathbf{R}_+ | Z_n = z, Z_n^-,\ldots,Z_1^-) \\ &\qquad = \mathbf{E}\left[\mathbf{E}\big[\mathbf{1}_{\{S_{n+1}\in\Phi_z^{-1}(B)\cap\mathbf{R}_+\}}\big| Z_n = z, \sigma(Z_i,S_i\,:\,0\leq i\leq n)\big]\,\Big| Z_n = z, Z_n^-,\ldots,Z_1^-\right] \\ &\qquad = \mathbf{E}\left[\mathcal{S}(z,\Phi_z^{-1}(B)\cap\mathbf{R}_+)\middle| Z_n = z, Z_n^-,\ldots,Z_1^-\right] \\ &\qquad = \mathcal{S}(z,\Phi_z^{-1}(B)\cap\mathbf{R}_+). \end{aligned}$$

As a consequence,

$$\begin{aligned} \mathbf{P}(Z_{n+1}^-\in B|\; Z_n^-,\ldots,Z_1^-) &= \int_E \mathbf{P}(Z_{n+1}^-\in B|Z_n = z, Z_n^-,\ldots,Z_1^-)Q(Z_n^-,\mathrm{d}z) \\ &= \int_E \mathcal{S}(z,\Phi_z^{-1}(B)\cap\mathbf{R}_+)Q(Z_n^-,\mathrm{d}z), \end{aligned}$$

which provides the result. $\square$

We focus on the ergodicity of (Z_n^-) by using Doeblin's condition. First, we consider the function $m : \overline{E} \to \mathbf{R}_+$ defined for any $z \in \overline{E}$ by

$$m(z) = \inf_{y \in \overline{E}} q(y, z). \tag{3}$$

In addition, μ denotes the measure on $(\overline{E}, \mathcal{B}(\overline{E}))$ defined by,

$$\forall B \in \mathcal{B}(\overline{E}),\ \mu(B) = \mu^0(B) + \mu^1(B), \tag{4}$$

where the measures μ^0 and μ^1 are given by

$$\begin{aligned} \mu^0(B) &= \int_E m(z) \left(\int_{\Phi_z^{-1}(B) \cap [0, t^+(z)[} f(z,s) \mathrm{d}s \right) \mathrm{d}z, \\ \mu^1(B) &= \int_E m(z) \mathbf{1}_{\Phi_z^{-1}(B)}(t^+(z)) G(z, t^+(z)) \mathrm{d}z. \end{aligned}$$

μ^0 may only charge the interior of the state space, while μ^1 may be strictly positive only on the boundary of E. By the expression of $\mathcal{R}$ (2) and the one of m (3), we have the straightforward inequality

$$\forall y \in \overline{E},\ \forall B \in \mathcal{B}(\overline{E}),\ \mathcal{R}(y, B) \geq \mu(B). \tag{5}$$

The first assumption that we impose concerns the measure μ.

Assumption 3.2. *We assume that the measure μ defined by (4) is a nondegenerate one. That is, $\mu/\mu(\overline{E})$ is a probability measure.*

Under this assumption, one may state that the Markov chain (Z_n^-) is ergodic.

Proposition 3.3. *We have the following results.*

- *The Markov chain (Z_n^-) is μ-irreducible, aperiodic and admits a unique invariant measure, which we denote by π.*
- *There exits $\rho > 1$ and $\kappa > 0$ such that,*

$$\forall n \geq 1,\ \sup_{\xi \in \overline{E}} \|\mathcal{R}^n(\xi, \cdot) - \pi\|_{TV} \leq \kappa \rho^{-n}, \tag{6}$$

  where $\|\cdot\|_{TV}$ stands for the total variation norm.
- *In addition, (Z_n^-) is positive Harris-recurrent.*

Proof. By definition and the inequality (5), (Z_n^-) is μ-irreducible and aperiodic (see [18, page 114]). Moreover, the transition kernel $\mathcal{R}$ obviously satisfies Doeblin's condition (see [18, page 396] for instance),

$$\mu(B) > \varepsilon \ \Rightarrow \mathcal{R}(y, B) > \varepsilon,$$

once again by (5). On the strength of Theorem 16.0.2 of [18], (Z_n^-) admits a unique invariant measure π since it is aperiodic and (6) holds. In addition, from Theorem 4.3.3 of [12], (Z_n^-) is positive Harris-recurrent. $\square$

Now, we shall impose some assumptions on the characteristics relative to the flow of the process. Under these new constraints, one may provide a more useful expression of $\mathcal{R}$. In the sequel, for $\xi \in E$, $t^-(\xi)$ denotes the deterministic exit time from E for the reverse flow,

$$t^-(\xi) = \sup\{t < 0 \,:\, \Phi_x(t) \in \partial E\},$$

with the usual convention $\sup \emptyset = -\infty$. Remark that $t^-(\xi)$ is a negative number.

Assumptions 3.4.

(i) The flow Φ is assumed to be $\mathcal{C}^1$-smooth. For any $(z,t)\in\mathbf{R}^d\times\mathbf{R}$, $D\Phi_z(t)$ is defined by

$$D\Phi_z(t) = \left|\det\left(\frac{\partial\Phi_x^{(i)}(t)}{\partial x_j}\right)_{1\leq i,j\leq d}\right|. \tag{7}$$

(ii) For any $t\in\mathbf{R}$, $\varphi_t:\mathbf{R}^d\to\mathbf{R}^d$, defined by $\varphi_t(x)=\Phi_x(t)$, is an injective application.

A useful expression of the transition kernel $\mathcal{R}$ is stated in the following proposition.

Proposition 3.5. *Let $y\in E$ and $B\in\mathcal{B}(\overline{E})$. We have*

$$\mathcal{R}(y,B) = \int_{B\cap E} r(y,z)\mathrm{d}z \quad + \ \mathcal{R}(y,B\cap\partial E), \tag{8}$$

where the conditional density function r is given by,

$$\forall z\in E,\ r(y,z) = \int_0^{-t^-(z)} q(y,\Phi_z(-s))f(\Phi_z(-s),s)D\Phi_z(-s)\mathrm{d}s. \tag{9}$$

Proof. Let $B\subset E$. First, we fix $t\geq 0$. We define the set A_t by

$$A_t = \{\Phi_\xi(-t)\ :\ \xi\in B\},$$

and we examine the function $\varphi_t : A_t\to B$ defined by $\varphi_t(x)=\Phi_x(t)$, $x\in A_t$. φ_t is a $\mathcal{C}^1$-smooth injective application (see Assumptions 3.4). Furthermore, for any $z\in B$,

$$\varphi_t(\Phi_z(-t)) = \Phi_{\Phi_z(-t)}(t) = z,$$

with $\Phi_z(-t)\in A_t$ by definition of A_t. Consequently, φ_t is a $\mathcal{C}^1$-one-to-one correspondence. The inverse function φ_t^{-1} is given by $\varphi_t^{-1}(x)=\Phi_x(-t)$, so it is $\mathcal{C}^1$-smooth too. Thus, φ_t is a $\mathcal{C}^1$-diffeomorphism from A_t into B, which allows us to consider it as a change of variable. In particular, this shows the relation

$$(z\in E,\, t\in\mathbf{R}_+, \Phi_z(t)\in B) \ \Leftrightarrow\ (t\in\mathbf{R}_+,\ z\in E,\ z\in A_t)\,. \tag{10}$$

Moreover, the Jacobian matrix $\mathbf{J}_{\varphi_t^{-1}}$ of the inverse function φ_t^{-1} satisfies,

$$\forall x\in\mathbf{R}^d,\ \mathbf{J}_{\varphi_t^{-1}}(x) = \left(\frac{\partial\Phi_x^{(i)}(-t)}{\partial x_j}\right)_{1\leq i,j\leq d}.$$

By (1) and (2), we have

$$\mathcal{R}(y,B) = \int_E\int_{\Phi_z^{-1}(B)\cap\mathbf{R}_+} q(y,t)f(z,t)\mathrm{d}t.$$

Together with (10), we obtain

$$\mathcal{R}(y,B) = \int_{\mathbf{R}_+}\left(\int_{A_t}\mathbf{1}_E(z)f(z,t)q(y,z)\mathrm{d}z\right)\mathrm{d}t.$$

By the change of variable φ_t, we have

$$\begin{aligned}\mathcal{R}(y,B) &= \int_{\mathbf{R}_+}\left(\int_B \mathbf{1}_E\left(\varphi_t^{-1}(\xi)\right)f\left(\varphi_t^{-1}(\xi),t\right)q\left(y,\varphi_t^{-1}(\xi)\right)\left|\det\mathbf{J}_{\varphi_t^{-1}}(\xi)\right|\mathrm{d}\xi\right)\mathrm{d}t\\ &= \int_{\mathbf{R}_+}\left(\int_B\mathbf{1}_E\left(\Phi_\xi(-t)\right)f\left(\Phi_\xi(-t),t\right)q\left(y,\Phi_\xi(-t)\right)D\Phi_\xi(-t)\mathrm{d}\xi\right)\mathrm{d}t,\end{aligned}$$

where $D\Phi_\xi$ is defined by (7). We remark that

$$(\Phi_\xi(-t) \in E) \Leftrightarrow (0 \leq t < -t^\star(\xi)),$$

so $\mathbf{1}_E(\Phi_\xi(-t)) = \mathbf{1}_{\{0\leq t<-t^\star(\xi)\}}$. By Fubini's theorem, this yields to the expected result. $\square$

In the light of this result, one may obtain the following one about the invariant distribution π of the Markov chain (Z_n^-): π admits a density with respect to the Lebesgue measure in the interior of the state space. In addition, one may exhibit a link between this density and r.

Corollary 3.6. *There exists a non-negative function p such that*

$$\forall B \in \mathcal{B}(\overline{E}),\ \pi(B) = \int_{B\cap E} p(x)\mathrm{d}x \ + \ \pi(B \cap \partial E). \tag{11}$$

In addition, p is given by the expression,

$$\forall x \in E,\ p(x) = \int_{\overline{E}} \pi(\mathrm{d}y) r(y,x). \tag{12}$$

Proof. μ is an irreducibility measure for $\mathcal{R}$. As a consequence and according to Theorem 4.2.2 of [18], the maximal irreducibility measure $\widetilde{\mu}$ is equivalent to the measure $\widetilde{\mu}'$ given for any $B \in \mathcal{B}(\overline{E})$, by

$$\widetilde{\mu}'(B) = \int_{\overline{E}} \mu(\mathrm{d}y) \sum_{n\geq 0} \mathcal{R}^n(y,B)\frac{1}{2^{n+1}}.$$

$\mathcal{R}$ admits a density on the interior E of the state space $\overline{E}$ of the Markov chain (Z_n^-) (see Proposition 3.5). As a consequence, this is the case for $\mathcal{R}^n$, too. Indeed, for any set B such that $\lambda_d(B \cap E) = 0$, we have

$$\mathcal{R}^n(y, B\cap E) = \int_{\overline{E}} \mathcal{R}^{n-1}(y,\mathrm{d}z)\mathcal{R}(z, B\cap E) = 0.$$

Therefore, $\widetilde{\mu}'(B\cap E) = 0$. Finally,

$$\lambda_d(B\cap E) = 0 \ \Rightarrow \ \widetilde{\mu}(B\cap E) = 0.$$

Since π and the maximal irreducibility measure $\widetilde{\mu}$ are equivalent, π admits a density on E. Now, we investigate the expression of this density. By Fubini's theorem, we have for any $B \subset E$,

$$\begin{aligned}
\pi(B) &= \int_{\overline{E}} \pi(\mathrm{d}y)\mathcal{R}(y,B) \\
&= \int_{\overline{E}} \pi(\mathrm{d}y) \int_B r(y,x)\mathrm{d}x \\
&= \int_B \left(\int_{\overline{E}} \pi(\mathrm{d}y) r(y,x) \right) \mathrm{d}x,
\end{aligned}$$

As a consequence, one may identify p with the function $\displaystyle\int_{\overline{E}} \pi(\mathrm{d}y) r(y,\cdot)$. $\square$

One shall see that the regularity of the conditional probability density function r is significant in all the sequel. We state that under additional assumptions, r is Lipschitz.

Assumptions 3.7. *We assume the following statements.*

(i) t^- *is a bounded and Lipschitz function, that is,*

$$\exists [t^-]_{Lip} > 0,\ \forall x, y \in E,\quad \left| t^-(x) - t^-(y) \right| \le [t^-]_{Lip} |x - y|.$$

(ii) The flow Φ *is Lipschitz, that is,*

$$\exists [\Phi]_{Lip} > 0,\ \forall x, y \in \mathbf{R}^d,\ \forall t \in \mathbf{R},\ \mid \Phi_x(t) - \Phi_y(t) \mid\ \le [\Phi]_{Lip} |x - y|.$$

(iii) f *is a bounded and Lipschitz function, that is,*

$$\exists [f]_{Lip} > 0,\ \forall x, y \in \mathbf{R}^d,\ \forall t \in \mathbf{R},\ |f(x,t) - f(y,t)|\ \le [f]_{Lip} |x - y|.$$

(iv) q *is a bounded and Lipschitz function, that is,*

$$\exists [q]_{Lip} > 0,\ \forall x, y, z \in \mathbf{R}^d,\ |q(x,y) - q(x,z)|\ \le [q]_{Lip} |y - z|.$$

(v) $D\Phi$ *is a bounded and Lipschitz function, that is,*

$$\exists [D\Phi]_{Lip} > 0,\ \forall x, y \in \mathbf{R}^d,\ \forall t \in \mathbf{R},\ |D\Phi_x(t) - D\Phi_y(t)|\ \le [D\Phi]_{Lip} |x - y|.$$

Proposition 3.8. r *is a bounded function. Furthermore, there exists a constant* $[r]_{Lip} > 0$ *such that, for any* $x \in \overline{E}$, $y \in E$ *and* $u \in \mathbf{R}^d$ *such that* $y + u \in E$*, we have*

$$|r(x, y + u) - r(x, y)| \le [r]_{Lip} |u|.$$

Proof. First, we have from (9),

$$\|r\|_\infty\ \le \|t^-\|_\infty\ \|q\|_\infty\ \|f\|_\infty \|D\Phi\|_\infty.$$

For the second point, we consider the function γ defined by,

$$\forall (y,t) \in \mathbf{R}^d \times \mathbf{R},\ \gamma(y,t) = q(x, \Phi_y(-t)) f(\Phi_y(-t), t) D\Phi_y(-t).$$

This function is Lipschitz as a compound and product of Lipschitz functions (see Assumptions 3.7). $[\gamma]_{Lip}$ stands for its Lipschitz constant, and we have

$$|\gamma(y,t) - \gamma(y+u,t)|\ \le [\gamma]_{Lip}\ |u|\ .$$

In addition, by (9), the function $r(x,y)$ is given by

$$r(x,y) = \int_0^{-t^-(y)} \gamma(y,s) \mathrm{d}s.$$

We suppose that $-t^-(y) \le -t^-(y+u)$ (recall that t^- is a negative function). We have

$$r(x, y+u) - r(x,y) = \int_0^{-t^-(y)} \left(\gamma(y+u,s) - \gamma(y,s)\right) \mathrm{d}s + \int_{-t^-(y)}^{-t^-(y+u)} \gamma(y+u,s) \mathrm{d}s.$$

As a consequence,

$$\begin{aligned} |r(x,y+u) - r(x,y)| \ &\le\ \int_0^{\|t^-\|_\infty} |\gamma(y+u,s) - \gamma(y,s)|\, \mathrm{d}s \\ &\quad +\ \|q\|_\infty \|f\|_\infty \|D\Phi\|_\infty \left| t^-(y) - t^-(y+u) \right| \\ &\le\ \|t^-\|_\infty [\gamma]_{Lip} |u| + [t^-]_{Lip}\ \|q\|_\infty \|f\|_\infty \|D\Phi\|_\infty |u|. \end{aligned}$$

The obtained inequality for $-t^-(y) > -t^-(y+u)$ is exactly the same one. This achieves the proof. □

3.2. **Estimation of p.** We propose a recursive nonparametric estimator of the function p given in the Corollary 3.6. In all the sequel, we consider a kernel function K which satisfies the following assumptions.

Assumptions 3.9. *The kernel $K : \mathbf{R}^d \to \mathbf{R}_+$ satisfies:*

(i) $\operatorname{supp} K \subset B(0_{\mathbf{R}^d}, \delta)$*, where $\delta > 0$. $B(x, r)$ stands for the open ball centered at x with radius r.*
(ii) K is a bounded function.

Under Assumptions 3.9, $\int_{\mathbf{R}^d} K^2(y)\mathrm{d}y$ is finite. τ^2 denotes this integral in the sequel. For all integer n, the recursive estimator $\widehat{p}_n$ of p that we propose is given for all $x \in E$ by

$$\widehat{p}_n(x) = \frac{1}{n} \sum_{j=1}^{n+1} \frac{1}{v_j^d} K\left(\frac{Z_j^- - x}{v_j}\right), \tag{13}$$

where the bandwith v_j satisfies

$$v_j = v_1 j^{-\alpha}, \text{ with } \alpha > 0.$$

Remark 3.10. *Let $x \in E$ and $j \geq 1$. Since the sequence (v_n) is decreasing, we have*

$$\operatorname{supp} K\left(\frac{\cdot - x}{v_j}\right) \subset \operatorname{supp} K\left(\frac{\cdot - x}{v_1}\right) \subset B(x, v_1\delta).$$

Thus, if $v_1\delta <$ $\operatorname{dist}(x, \partial E)$, we have

$$\operatorname{supp} K\left(\frac{\cdot - x}{v_j}\right) \subset E.$$

In the following proposition, we establish the pointwise asymptotic consistency of $\widehat{p}_n$.

Proposition 3.11. *Let $x \in E$. One chooses v_1 such that $v_1\delta < \operatorname{dist}(x, \partial E)$ and α such that $\alpha d < 1$. Then,*

$$\widehat{p}_n(x) \xrightarrow{a.s.} p(x),$$

when n goes to infinity.

Proof. By the expression of $p(x)$ given by (12), the difference $\widehat{p}_n(x) - p(x)$ may be written in the following way,

$$\begin{aligned}
\widehat{p}_n(x) - p(x) &= \frac{1}{n} \sum_{j=1}^{n+1} \frac{1}{v_j^d} K\left(\frac{Z_j^- - x}{v_j}\right) - \int_{\overline{E}} r(u, x)\pi(\mathrm{d}u) \\
&= \frac{1}{n v_1^d} K\left(\frac{Z_1^- - x}{v_1}\right) + \frac{1}{n} M_n + R_n^{(1)} + R_n^{(2)},
\end{aligned} \tag{14}$$

where M_n, $R_n^{(1)}$ and $R_n^{(2)}$ are given by

$$\begin{aligned} M_n &= \sum_{j=1}^{n}\left[\frac{1}{v_{j+1}^d}K\left(\frac{Z_{j+1}^- - x}{v_{j+1}}\right) - \int_{\mathbf{R}^d} r(Z_j^-, x+yv_{j+1})K(y)\mathrm{d}y\right], \\ R_n^{(1)} &= \frac{1}{n}\sum_{j=1}^{n}\int_{\mathbf{R}^d}\left[r(Z_j^-, x+yv_{j+1}) - r(Z_j^-, x)\right]K(y)\mathrm{d}y, \\ R_n^{(2)} &= \frac{1}{n}\sum_{j=1}^{n} r(Z_j^-, x) - \int_{\overline{E}} r(u,x)\pi(\mathrm{d}u). \end{aligned}$$

The dependency on x is implicit. In (14), the first term clearly tends to 0 as n goes to infinity. The sequel of the proof is divided into three parts: in the first one, we show that $R_n^{(2)}$ tends to 0 by the ergodic theorem. In the second one, we focus on $R_n^{(1)}$ and we prove that this term goes to 0. Finally, we state that M_n/n tends to 0 by using the second law of large numbers for martingales.

Recall that the Markov chain (Z_n^-) is positive Harris-recurrent with invariant measure π, according to Proposition 3.3. Thus, one may apply the ergodic theorem (see for instance Theorem 17.1.7 of [18]) and we obtain that $R_n^{(2)}$ almost surely tends to 0. For $R_n^{(1)}$, we have

$$\begin{aligned} \left|R_n^{(1)}\right| &\leq \frac{1}{n}\sum_{j=1}^{n}\int_{\mathbf{R}^d}\left|r(Z_j^-, x+yv_{j+1}) - r(Z_j^-, x)\right|K(y)\mathrm{d}y \\ &\leq \frac{1}{n}\sum_{j=1}^{n}\int_{\mathbf{R}^d}[r]_{Lip}|y|v_{j+1}K(y)\mathrm{d}y \\ &\leq \frac{1}{n}\sum_{j=1}^{n} v_{j+1}\left(\int_{\mathbf{R}^d}|y|K(y)\mathrm{d}y\right)[r]_{Lip}. \end{aligned}$$

This upper bound tends to 0 by Cesaro's lemma because the limit of the sequence (v_n) is 0. Therefore, $R_n^{(1)}$ goes to 0 as n tends to infinity. Finally, we investigate the term M_n/n. First, we show that the process (M_n) is a discrete-time martingale with respect to the filtration $(\mathcal{F}_n)$ defined by,

$$\forall n \geq 1,\ \mathcal{F}_n = \sigma(Z_1^-, \ldots, Z_n^-, Z_{n+1}^-).$$

We have

$$\mathbf{E}\left[M_n|\mathcal{F}_{n-1}\right] = M_{n-1} + \mathbf{E}\left[\frac{1}{v_{n+1}^d}K\left(\frac{Z_{n+1}^- - x}{v_{n+1}}\right)\middle| Z_n^-\right] - \int_{\mathbf{R}^d} r(Z_n^-, x+yv_{n+1})K(y)\mathrm{d}y.$$

Thus, we only have to prove that

$$\mathbf{E}\left[\frac{1}{v_{n+1}^d}K\left(\frac{Z_{n+1}^- - x}{v_{n+1}}\right)\middle| Z_n^-\right] = \int_{\mathbf{R}^d} r(Z_n^-, x+yv_{n+1})K(y)\mathrm{d}y. \tag{15}$$

We have

$$\mathbf{E}\left[\frac{1}{v_{n+1}^d}K\left(\frac{Z_{n+1}^- - x}{v_{n+1}}\right)\middle| Z_n^-\right] = \frac{1}{v_{n+1}^d}\int_{\overline{E}} K\left(\frac{u-x}{v_{n+1}}\right)\mathcal{R}(Z_n^-, \mathrm{d}u).$$

By the assumption on v_1 and Remark 3.10,

$$\begin{aligned}\int_{\overline{E}} K\left(\frac{u-x}{v_{n+1}}\right)\mathcal{R}(Z_n^-,\mathrm{d}u) &= \int_E K\left(\frac{u-x}{v_{n+1}}\right)\mathcal{R}(Z_n^-,\mathrm{d}u)\\ &= \int_E K\left(\frac{u-x}{v_{n+1}}\right) r(Z_n^-,u)\mathrm{d}u,\end{aligned}$$

by (8). Finally, the change of variable $u = yv_{n+1} + x$ states (15). Thus, (M_n) is a martingale. We shall study the asymptotic behavior of its predictable quadratic variation $< M >$. A straightforward calculus leads to

$$\begin{aligned}(M_n - M_{n-1})^2 &= \frac{1}{v_{n+1}^{2d}}K^2\left(\frac{Z_{n+1}^- - x}{v_{n+1}}\right) + \left[\int_E r(Z_n^-, x+yv_{n+1})K(y)\mathrm{d}y\right]^2\\ &\quad -\frac{2}{v_{n+1}^d}K\left(\frac{Z_{n+1}^- - x}{v_{n+1}}\right)\int_E r(Z_n^-, x+yv_{n+1})K(y)\mathrm{d}y.\end{aligned}$$

Using the method used to show (15), we deduce that

$$\begin{aligned}\mathbf{E}\left[(M_n - M_{n-1})^2|\mathcal{F}_n\right] &= \frac{1}{v_{n+1}^d}\int_E K^2(y) r(Z_n^-, x+yv_{n+1})\mathrm{d}y\\ &\quad -\left[\int_E r(Z_n^-, x+yv_{n+1})K(y)\mathrm{d}y\right]^2.\end{aligned}$$

As a consequence,

$$< M >_n \leq \sum_{j=1}^n \left(\frac{1}{v_{j+1}^d}\ \|r\|_\infty \tau^2 + \|r\|_\infty\right) \sim \mathrm{cst}\, n^{\alpha d+1},$$

as n tends to infinity. By the second law of large numbers for martingales (see Theorem 1.3.15 of [9]), we have

$$M_n^2 = \mathcal{O}\left(< M >_n \ln(< M >_n)^{1+\gamma}\right), \quad \text{with } \gamma > 0.$$

As a consequence,

$$\frac{M_n}{n} = \mathcal{O}\left(\sqrt{n^{\alpha d-1}\ln(n^{\alpha d+1})^{1+\gamma}}\right).$$

Thus, M_n/n almost surely tends to 0 as n goes to infinity if $\alpha d < 1$. This achieves the proof. □

4. Estimation of the invariant distribution of (Z_n^-, Z_n)

In this section, we state that the Markov chain (Z_n^-, Z_n) admits a unique invariant measure. In addition, we are interested in the recursive estimation of this measure.

4.1. **Some properties of (Z_n^-, Z_n).** In this part, we focus on the asymptotic behavior of the chain (Z_n^-, Z_n). In all the sequel, η_n (respectively π_n) denotes the distribution of

(Z_n^-, Z_n) (resp. Z_n^-) for all integer n. We have these straightforward relations between η_n, Q or q and π_n,

$$\begin{aligned} \eta_n(A \times B) &= \int_A Q(z, B)\pi_n(\mathrm{d}z) \\ &= \int_{A\times B} q(z, y)\pi_n(\mathrm{d}z)\mathrm{d}y. \end{aligned} \tag{16}$$

Lemma 4.1. *We have*

$$\lim_{n\to+\infty} \|\eta_n - \eta\|_{TV} = 0,$$

where the limit distribution η is defined for all $A \times B \in \mathcal{B}(\overline{E} \times E)$ by

$$\eta(A \times B) = \int_{A\times B} q(z, y)\pi(\mathrm{d}z)\mathrm{d}y. \tag{17}$$

Proof. Let g be a measurable function bounded by 1. By virtue of Fubini's theorem, we have

$$\left| \int_{\overline{E}\times E} g(x, y) \left(\eta_n(\mathrm{d}x \times \mathrm{d}y) - \eta(\mathrm{d}x \times \mathrm{d}y)\right)\right| \leq \left| \int_{\overline{E}} \left(\pi_n(\mathrm{d}x) - \pi(\mathrm{d}x)\right) \int_E g(x, y)q(x, y)\mathrm{d}y\right|,$$

from (16) and (17). Thus,

$$\left| \int_{\overline{E}\times E} g(x, y) \left(\eta_n(\mathrm{d}x \times \mathrm{d}y) - \eta(\mathrm{d}x \times \mathrm{d}y)\right)\right| \leq \left| \int_{\overline{E}} \widetilde{g}(x)(\pi_n(\mathrm{d}x) - \pi(\mathrm{d}x))\right|,$$

where the function $\widetilde{g} : x \mapsto \int_E g(x, y)q(x, y)\mathrm{d}y$ is bounded by 1 since g is bounded by 1 and q is the conditional density associated with the Markov kernel Q. As a consequence,

$$\|\eta_n - \eta\|_{TV} \leq \|\pi_n - \pi\|_{TV}. \tag{18}$$

One obtains the expected limit from (6). □

In addition, one may prove that η admits a density on $E \times E$.

Lemma 4.2. *There exists a positive function h such that*

$$\eta(A \times B) = \int_{A\times B} h(x, y)\mathrm{d}x\, \mathrm{d}y,$$

for any $A \times B \in \mathcal{B}(\overline{E} \times E)$ with $A \subset E$. In addition, h is given for all $x, y \in E$ by

$$h(x, y) = p(x)q(x, y). \tag{19}$$

Proof. From (17), we have

$$\begin{aligned} \eta(A \times B) &= \int_{A\times B} q(z, y)\pi(\mathrm{d}z)\mathrm{d}y \\ &= \int_{A\times B} q(z, y)p(z)\mathrm{d}z\, \mathrm{d}y, \end{aligned}$$

by (11) and because $A \subset E$. This achieves the proof. □

4.2. **Estimation of h.** We propose to estimate the function h by the recursive nonparametric estimator $\widehat{h}_n$ given, for any $(x,y)\in E^2$, by

$$\widehat{h}_n(x,y)=\frac{1}{n}\sum_{j=1}^{n+1}\frac{1}{w_j^{2d}}K\left(\frac{Z_j^- - x}{w_j}\right)K\left(\frac{Z_j-y}{w_j}\right), \tag{20}$$

where the bandwith w_j is given by

$$w_j = w_1 j^{-\beta},\ \text{ with } \beta>0.$$

The kernel function K is assumed to satisfy Assumptions 3.9. In the sequel, we are interested in the pointwise convergence of the estimator at a point $(x,y)\in E^2$. We assume that w_1 is such that $w_1\delta<\mathrm{dist}(x,\partial E)$, where δ is the radius of the open ball which contains the support of the kernel function K. In this case, Remark 3.10 is still valid, and we have the following inclusions, for any integer j,

$$\operatorname{supp} K\left(\frac{\cdot - x}{w_j}\right)\subset B(x,w_1\delta)\subset E. \tag{21}$$

Our main objective is to state in Proposition 4.7 that $\widehat{h}_n(x,y)$ almost surely converges to $h(x,y)$. First, we show that this estimator is asymptotically unbiased (see Proposition 4.4).

We state some new properties of the distribution measures π_n and π. Let us recall that π_n is the law of Z_n^-, while π is the invariant measure of the Markov chain (Z_n^-).

Lemma 4.3. *We have the following statements.*

- *For any integer n, π_n admits a density function p_n on E.*
- *p_n is bounded by $\|r\|_\infty$ and is an $[r]_{Lip}$-Lipschitz function.*
- *p is Lipschitz.*
- *For any integer n, we have*

$$\sup_{x\in E}|p_n(x)-p(x)|\leq \|r\|_\infty\kappa\rho^{-(n-1)}. \tag{22}$$

Proof. For the first point, let $B\in\mathcal{B}(\overline{E})$ with $B\subset E$. We have

$$\begin{aligned}\pi_n(B) &= \int_{\overline{E}}\int_B \mathcal{R}(\xi,\mathrm{d}y)\pi_{n-1}(\mathrm{d}\xi)\\ &= \int_B\int_{\overline{E}} r(\xi,y)\pi_{n-1}(\mathrm{d}\xi)\mathrm{d}y,\end{aligned}$$

where r is the conditional density associated with the kernel $\mathcal{R}$ (see (8)). Thus, one may identify

$$p_n(y)=\int_{\overline{E}} r(\xi,y)\pi_{n-1}(\mathrm{d}\xi). \tag{23}$$

For the second assertion, we have stated in Proposition 3.8 that r is a bounded function. As a consequence,

$$|p_n(y)| \ \leq\ \|r\|_\infty\pi_{n-1}(\overline{E})\ =\ \|r\|_\infty.$$

In addition, since r is Lipschitz,

$$\begin{aligned}|p_n(y)-p_n(z)| &\leq \int_{\overline{E}} |r(\xi,y)-r(\xi,z)|\,\pi_{n-1}(\mathrm{d}\xi)\\ &\leq [r]_{Lip}|y-z|\pi_{n-1}(\overline{E}) \quad = \quad [r]_{Lip}|y-z|.\end{aligned}$$

For the third point, p is Lipschitz for the same reason than p_n since p satisfies (12). Finally, for the last point, we have, by (12) and (23),

$$\begin{aligned}|p_n(x)-p(x)| &\leq \int_{\overline{E}} r(y,x)\,|\pi_{n-1}(\mathrm{d}y)-\pi(\mathrm{d}y)|\\ &\leq \|r\|_\infty\,\|\pi_{n-1}-\pi_n\|_{TV}\\ &\leq \|r\|_\infty \kappa\rho^{-(n-1)},\end{aligned}$$

by (6). This achieves the proof. □

Now, one may state that $\widehat{h}_n(x,y)$ is an asymptotically unbiased estimator of $h(x,y)$.

Proposition 4.4. *When n goes to infinity,*

$$\mathbf{E}\left[\widehat{h}_n(x,y)\right] \to h(x,y).$$

Proof. We only state that

$$\mathbf{E}\left[\widehat{h}_n(x,y)\right] - \frac{n+1}{n}h(x,y)$$

tends to 0. We have

$$\begin{aligned}\mathbf{E}\left[\widehat{h}_n(x,y)\right] - \frac{n+1}{n}h(x,y) &= \frac{1}{n}\sum_{j=1}^{n+1}\Bigg[\int_{\overline{E}\times E}\frac{1}{w_j^{2d}}K\left(\frac{u-x}{w_j}\right)K\left(\frac{v-y}{w_j}\right)\pi_j(\mathrm{d}u)q(u,v)\mathrm{d}v\\ &\qquad - \int_{E\times E} h(x,y)K(u)K(v)\mathrm{d}u\mathrm{d}v\Bigg].\end{aligned}$$

Thanks to (21), one may replace $\overline{E}$ by E in the first integral. As a consequence, one may replace $\pi_j(\mathrm{d}u)$ by $p_j(u)\mathrm{d}u$ (see Lemma 4.3). Together with (19) and a change of variables, we obtain

$$\begin{aligned}&\mathbf{E}\left[\widehat{h}_n(x,y)\right] - \frac{n+1}{n}h(x,y)\\ &= \frac{1}{n}\sum_{j=1}^{n+1}\int_{E\times E} K(u)K(v)\Big(p_j(x+uw_j)q(x+uw_j,y+vw_j)-p(x)q(x,y)\Big)\mathrm{d}u\,\mathrm{d}v.\end{aligned}$$

Furthermore, since p_j is $[r]_{Lip}$-Lipschitz and bounded by $\|r\|_\infty$ in the light of Lemma 4.3, an elementary calculus leads to

$$\begin{aligned}&\big|p_j(x+uw_j)\,q(x+uw_j,y+vw_j)-p(x)q(x,y)\big|\\ &\qquad \leq \|r\|_\infty[q]_{Lip}|v|w_j + \|q\|_\infty[r]_{Lip}|u|w_j + \|q\|_\infty\,|p_j(x)-p(x)|\\ &\qquad \leq w_j\left(\|r\|_\infty[q]_{Lip}|v| + \|q\|_\infty[r]_{Lip}|u|\right) + \|q\|_\infty\|r\|_\infty\kappa\rho^{-(j-1)},\end{aligned}$$

with (22). Finally, we obtain

$$\begin{aligned}
&\left|\mathbf{E}\left[\widehat{h}_n(x,y)\right] - \frac{n+1}{n}h(x,y)\right| \\
&\quad\leq \frac{(\|r\|_\infty[q]_{Lip} + \|q\|_\infty[r]_{Lip})\int_E K(u)|u|\mathrm{d}u}{n}\sum_{j=1}^{n+1} w_j + \frac{\|r\|_\infty\|q\|_\infty\kappa}{n}\sum_{j=1}^{n+1}\rho^{-(j-1)} \\
&\quad\leq \frac{(\|r\|_\infty[q]_{Lip} + \|q\|_\infty[r]_{Lip})\int_E K(u)|u|\mathrm{d}u}{n}\sum_{j=1}^{n+1} w_j + \frac{\|r\|_\infty\|q\|_\infty\kappa}{n(1-\rho^{-1})},
\end{aligned}$$

which tends to 0 by Cesaro's lemma. □

In the following, we are interested in some properties of the discrete-time process

$$(A_n) = \left(\frac{1}{w_n^{2d}}K\left(\frac{Z_n^- - x}{w_n}\right)K\left(\frac{Z_n - y}{w_n}\right)\right), \tag{24}$$

which naturally appears in the study of the estimator $\widehat{h}_n(x,y)$. In particular, we propose to investigate its autocovariance function. On the strength of this result, we will establish the asymptotic behavior of the variance of $\widehat{h}_n(x,y)$.

Proposition 4.5. *There exist two constants B and $b > 1$ such that, for any integers $n \geq k$,*

$$\left|\mathbf{C}\mathrm{ov}(A_k, A_n)\right| \leq \frac{\|K\|_\infty^4 B}{w_n^{4d}} b^{k-n}\left(1 + b^{-k}\right). \tag{25}$$

In particular, one obtains by taking $k = n$ and using that $b^{-n} \leq 1$,

$$\mathbf{V}\mathrm{ar}(A_n) \leq \frac{2\|K\|_\infty^4 B}{w_n^{4d}}. \tag{26}$$

Proof. We have

$$\begin{aligned}
&\mathbf{C}\mathrm{ov}(A_k, A_n) \\
&= \frac{\|K\|_\infty^4}{w_n^{2d}w_k^{2d}}\mathbf{C}\mathrm{ov}\left(\frac{1}{\|K\|_\infty^2}K\left(\frac{Z_k^- - x}{w_k}\right)K\left(\frac{Z_k - y}{w_k}\right), \frac{1}{\|K\|_\infty^2}K\left(\frac{Z_n^- - x}{w_n}\right)K\left(\frac{Z_n - y}{w_n}\right)\right),
\end{aligned}$$

where both the components in the covariance are bounded by 1. We apply Theorem 16.1.5 of [18] with $V = 1$, $\Phi = (Z^-, Z)$,

$$g = \frac{1}{\|K\|_\infty^2}K\left(\frac{\cdot - x}{w_n}\right)K\left(\frac{\cdot - y}{w_n}\right) \quad \text{and} \quad h = \frac{1}{\|K\|_\infty^2}K\left(\frac{\cdot - x}{w_k}\right)K\left(\frac{\cdot - y}{w_k}\right).$$

The conditions of the theorem are satisfied by (6) and (18). We obtain

$$|\mathbf{C}\mathrm{ov}(A_k, A_n)| \leq \frac{\|K\|_\infty^4}{w_n^{2d}w_k^{2d}} B b^{k-n}\left(1 + b^{-k}\right).$$

Together with $1/w_k^{2d} \leq 1/w_n^{2d}$, this shows (25). □

In the following result, we give a bound for the variance of $\widehat{h}_n(x,y)$. It is a corollary of Proposition 4.5.

Corollary 4.6. *Let n be an integer. We have*

$$\mathbf{V}\mathrm{ar}\left(\widehat{h}_n(x,y)\right) \leq \frac{8}{n w_{n+1}^{4d}} \frac{\|K\|_\infty^4 B}{1-b^{-1}}. \tag{27}$$

As a consequence, this variance goes to 0 *when* $4\beta d < 1$ *(recall that* $w_n = w_1 n^{-\beta}$*).*

Proof. This inequality is a consequence of (25) stated in Proposition 4.5. Indeed, in light of the expressions of A_n (24) and $\widehat{h}_n(x,y)$ (20), we have

$$\begin{aligned}\mathbf{V}\mathrm{ar}\left(\widehat{h}_n(x,y)\right) &= \frac{2}{n^2}\sum_{k=1}^{n+1}\sum_{l=k}^{n+1}\mathbf{C}\mathrm{ov}(A_l,A_k)\\ &\leq \frac{2}{n^2}\|K\|_\infty^4 B \sum_{k=1}^{n+1}\sum_{l=k}^{n+1} b^{k-l}\left(1+b^{-k}\right)w_l^{-4d}.\end{aligned}$$

Using that $b^{-k}\leq 1$ and $w_l^{-4d}\leq w_{n+1}^{-4d}$,

$$\begin{aligned}\mathbf{V}\mathrm{ar}\left(\widehat{h}_n(x,y)\right) &\leq \frac{4}{n^2 w_{n+1}^{4d}}\|K\|_\infty^4 B\sum_{k=1}^{n+1} b^k \sum_{l=k}^{n+1} b^{-l}\\ &\leq \frac{4}{n^2 w_{n+1}^{4d}}\|K\|_\infty^4\ B\sum_{k=1}^{n+1} b^k \frac{b^{-k}}{1-b^{-1}}\\ &\leq \frac{8}{n w_{n+1}^{4d}}\frac{\|K\|_\infty^4 B}{1-b^{-1}},\end{aligned}$$

with $(n+1)/n \leq 2$. □

Now, one may state the consistency of our estimator of $h(x,y)$.

Proposition 4.7. *Let $(x,y)\in E^2$. One chooses $w_1\delta < \mathrm{dist}(x,\partial E)$ and $8\beta d < 1$. Then,*

$$\widehat{h}_n(x,y) \xrightarrow{a.s.} h(x,y),$$

when n goes to infinity.

Proof. According to Proposition 4.4, we only have to prove that

$$Y_n = \left(\widehat{h}_n(x,y) - \mathbf{E}\left[\widehat{h}_n(x,y)\right]\right)^2 \tag{28}$$

almost surely converges to 0. In the sequel of the proof, we establish that there exists a random variable Y such that $Y_n \xrightarrow{a.s.} Y$. Since the sequence (Y_n) tends to 0 in $\mathbf{L}^1$ (remark that $\mathbf{E}[Y_n] = \mathbf{V}\mathrm{ar}\left(\widehat{h}_n(x,y)\right)$, together with Corollary 4.6), $Y = 0$ *a.s.* and it induces the expected result. In order to show the almost sure convergence of the sequence (Y_n), we use Van Ryzin's lemma (see [21]). In light of this result, if the sequence (Y_n) satisfies the following conditions,

(i) $Y_n \geq 0$ *a.s.*,
(ii) $\mathbf{E}[Y_1] < +\infty$,
(iii) $\mathbf{E}\left[Y_{n+1}|\mathcal{S}_n\right] \leq Y_n + Y_n'$ *a.s.*, where $\mathcal{S}_n = \sigma(Y_1,\ldots,Y_n)$ and Y_n' is $\mathcal{S}_n$-measurable,
(iv) $\sum_{n\geq 1}\mathbf{E}\left[|Y_n'|\right] < +\infty$,

then $Y_n \xrightarrow{a.s.} Y$. In our context, points (i) and (ii) are unquestionably satisfied. By (20), (24) and (28), we have

$$\frac{1}{n}\sum_{k=1}^{n+1}\big(A_k - \mathbf{E}[A_k]\big) = \sqrt{Y_n}.$$

Consequently, we have the recurrence relation,

$$\sqrt{Y_n} = \frac{A_{n+1} - \mathbf{E}[A_{n+1}] + (n-1)\sqrt{Y_{n-1}}}{n}.$$

By squaring, we obtain

$$\begin{aligned} Y_n &= \left(\frac{n-1}{n}\right)^2 Y_{n-1} + \frac{\big(A_{n+1} - \mathbf{E}[A_{n+1}]\big)^2 + 2(n-1)\sqrt{Y_{n-1}}\big(A_{n+1} - \mathbf{E}[A_{n+1}]\big)}{n^2} \\ &\leq Y_{n-1} + \frac{\big(A_{n+1} - \mathbf{E}[A_{n+1}]\big)^2 + 2(n-1)\sqrt{Y_{n-1}}\big(A_{n+1} - \mathbf{E}[A_{n+1}]\big)}{n^2}. \end{aligned}$$

Finally, $\mathbf{E}[Y_n|\mathcal{S}_{n-1}] \leq Y_{n-1} + Y'_{n-1}$, where

$$Y'_{n-1} = \frac{1}{n^2}\mathbf{E}\left[\big(A_{n+1} - \mathbf{E}[A_{n+1}]\big)^2 + 2(n-1)\sqrt{Y_{n-1}}\big(A_{n+1} - \mathbf{E}[A_{n+1}]\big)\Big|\, \mathcal{S}_{n-1}\right]. \tag{29}$$

Thus, (iii) is checked. Ultimately, we have to verify (iv). By (29) and Cauchy-Schwarz inequality,

$$\begin{aligned} \mathbf{E}\left[|Y'_{n-1}|\right] &\leq \frac{1}{n^2}\mathbf{V}\mathrm{ar}(A_{n+1}) + \frac{2(n-1)}{n^2}\mathbf{E}\left[\sqrt{Y_{n-1}}\,\big|A_{n+1} - \mathbf{E}[A_{n+1}]\big|\right] \\ &\leq \frac{1}{n^2}\mathbf{V}\mathrm{ar}(A_{n+1}) + \frac{2}{n}\sqrt{\mathbf{E}[Y_{n-1}]\mathbf{V}\mathrm{ar}(A_{n+1})} \\ &\leq \frac{1}{n^2}\mathbf{V}\mathrm{ar}(A_{n+1}) + \frac{2}{n}\sqrt{\mathbf{V}\mathrm{ar}\left(\widehat{h}_{n-1}(x,y)\right)\mathbf{V}\mathrm{ar}(A_{n+1})}. \end{aligned}$$

Thus, by (26) and (27), there exist two numbers c_1 and c_2 such that

$$\begin{aligned} \mathbf{E}\left[|Y'_{n-1}|\right] &\leq \frac{c_1}{n^2 w_n^{4d}} + \frac{c_2}{n^{3/2} w_n^{4d}} \\ &\leq \frac{c_1}{w_1 n^{2-4\beta d}} + \frac{c_2}{w_1 n^{3/2-4\beta d}}. \end{aligned}$$

As a consequence, $\sum \mathbf{E}[|Y'_n|]$ is a convergent series for $8\beta d < 1$. □

Our main convergence result is stated in the following theorem. The recursive estimator of $q(x,y)$ that we consider may be written as follows,

$$\widehat{q}_n(x,y) = \frac{\displaystyle\sum_{j=1}^{n+1} \frac{1}{w_j^{2d}} K\left(\frac{Z_j^- - x}{w_j}\right) K\left(\frac{Z_j - y}{w_j}\right)}{\displaystyle\sum_{j=1}^{n+1} \frac{1}{v_j^d} K\left(\frac{Z_j^- - x}{v_j}\right)},$$

where the couple (x,y) is in E^2, $v_j = v_1 j^{-\alpha}$, $w_j = w_1 j^{-\beta}$, with $\alpha, \beta > 0$, and K is a kernel function satisfying Assumptions 3.9.

Theorem 4.8. *Let us choose* v_1 *and* w_1 *such that* $\max(v_1, w_1)\delta < \mathrm{dist}(x, \partial E)$*. If* $p(x) > 0$*,* $\alpha d < 1$ *and* $8\beta d < 1$*, then, when* n *goes to infinity,*

$$\widehat{q}_n(x, y) \xrightarrow{a.s.} q(x, y).$$

Proof. In light of (19), if $p(x) > 0$, one may write $q(x, y) = h(x, y)/p(x)$. In addition, $\widehat{q}_n(x, y)$ is defined by the ratio $\widehat{h}_n(x, y)/\widehat{p}_n(x)$ (see (20) for the expression of $\widehat{h}_n(x, y)$ and (13) for the one of $\widehat{p}_n(x)$), where $\widehat{h}_n(x, y)$ (respectively $\widehat{p}_n(x)$) estimates $h(x, y)$ (resp. $p(x)$). In such a case, the result is a corollary of Propositions 3.11 and 4.7. □

5. Simulation study

The goal of this section is to illustrate the asymptotic behavior of our recursive estimator via numerical experiments in the one-dimensional case. More precisely, we investigate numerical simulations for a quite simple example in Subsection 5.1, and for the well-known TCP window size process in Subsection 5.2.

5.1. **First numerical experiments.**

We consider a PDMP (X_t) defined on the state space $E =]0, 1[$, and we assume that the process starts from the midpoint of E, $X_0 = 0.5$. The main characteristics of (X_t) are defined as follows. The flow Φ satisfies $\Phi_x(t) = x + t$ for any $x \in E$ and $t \in \mathbf{R}$. In addition, the jump rate λ is chosen constant equal to 10. Finally, for $x \in E$, $Q(x, \cdot)$ is the truncated exponential distribution with mean $1/(10 + x)$. All the conditions assumed in the present paper are obviously satisfied for this example. In particular, the transition kernel Q admits a density q with respect to the Lebesgue measure.

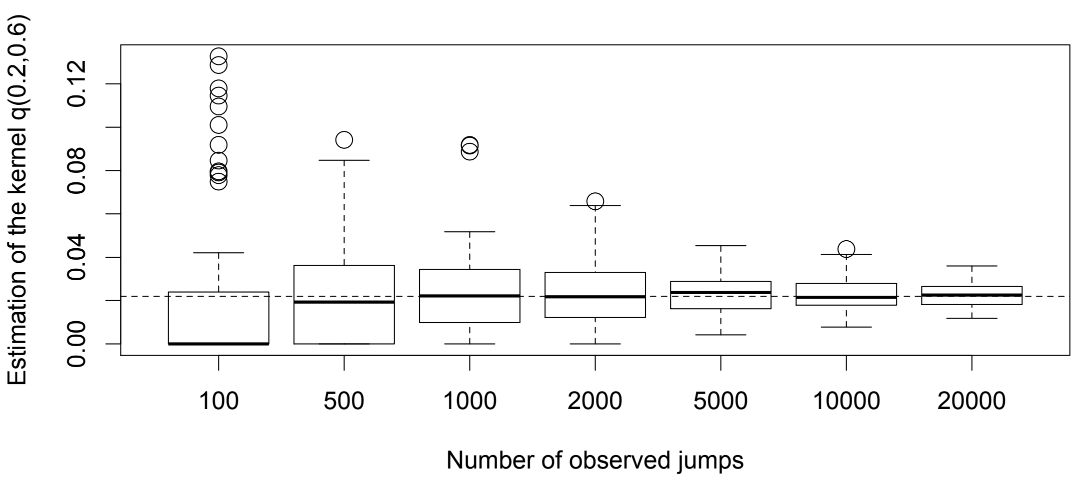

FIGURE 1. Boxplots, over 100 replicates, of the estimation of $q(x, y)$, with $x = 0.2$ and $y = 0.6$, from different numbers of observed jumps. When $\widehat{p}_n(x) = 0$, we replace $\widehat{q}_n(x, y)$ by 0. As a consequence, they are lots of zeros over the 100 replicates when n is too small.

In the sequel, we focus our attention on the nonparametric estimation of the conditional density q associated with Q. For any couple $(x,y) \in E^2$, the function of interest $q(x,y)$ satisfies

$$q(x,y) = \frac{1}{K_x}(10+x)\exp\left(-(10+x)y\right),$$

where K_x stands for the normalizing constant. We choose the Epanechnikov kernel K given, for any $x \in \mathbf{R}$, by

$$K(x) = \frac{3}{4}(1-x^2)\mathbf{1}_{\{|x|\leq 1\}},$$

which satisfies all the assumptions imposed in the theoretical part of the paper. Finally, we select the constant parameters α and β such that

$$\alpha = \beta = \frac{1}{8} - 10^{-2}.$$

Figure 1 presents the empirical distribution over 100 replicates of the estimate $\widehat{q}_n(x,y)$, with $x = 0.2$ and $y = 0.6$, from different numbers of observed jumps. On small-sampled sizes, our procedure is quite unfulfilling. However, for n large enough, our method succeeds in the estimation of the quantity of interest, especially when n is greater than 10000. In addition, the complete curve $q(x,y)$, with $x = 0.5$ and $0.1 \leq y \leq 0.9$, and its estimate from 20000 jumps are very close (see Figure 2). The estimation is provided only on the interval $[0.1, 0.9] \subset E$, since the almost sure convergence of $\widehat{q}_n$ has been stated on the interior of the state space.

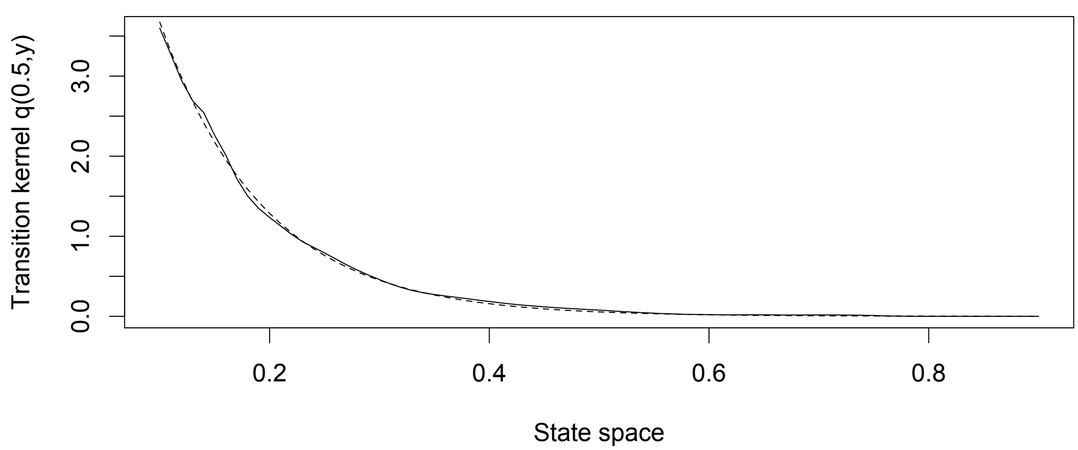

FIGURE 2. Estimation of $q(x,y)$, $x = 0.5$, for $0.1 \leq y \leq 0.9$ from 20000 observed jumps. The real function is drawn in dashed line, while its estimate is in solid line.

5.2. **The TCP window size process.** The TCP window size process naturally appears in the modeling of the famous Transmission Control Protocol used for data transmission over the Internet. For this special case of PDMP, the sample paths are piecewise-linear and the whole randomness of the dynamics comes from the jump times. The motion of

the TCP window size process (X_t) on $E =]0, +\infty[$ is governed by its features Φ, λ and Q. The flow Φ is linear, that is, $\Phi_x(t) = x + t$ for any $x \in E$ and $t \in \mathbf{R}$. The jump rate λ satisfies $\lambda(x) = x$. Finally, the transition kernel Q is given, for any couple $(x, y) \in E^2$, by $Q(x, \mathrm{d}y) = \delta_{\{x/2\}}(\mathrm{d}y)$. One may refer the interested reader to [4] for a study on the long time behavior of this process.

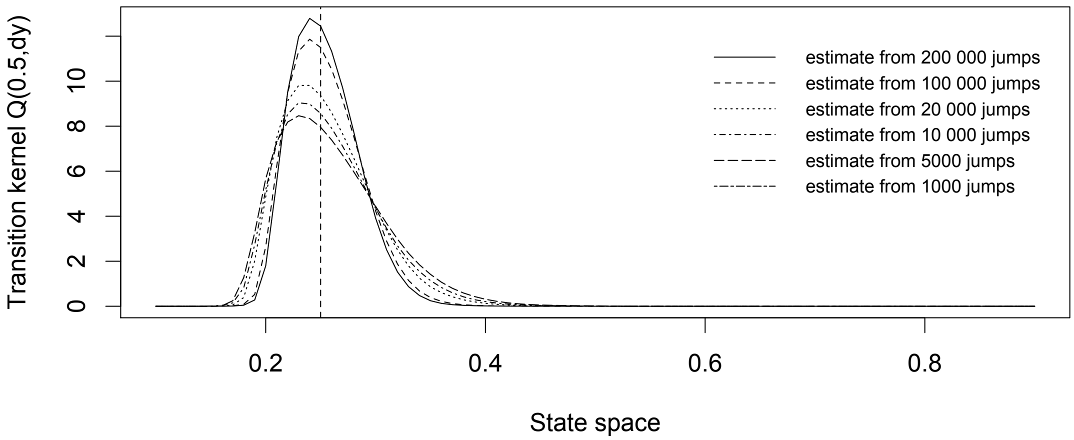

FIGURE 3. Estimation of $Q(x, \mathrm{d}y) = \delta_{\{x/2\}}(\mathrm{d}y)$, with $x = 0.5$, from different numbers of observed jumps. The vertical line stands for the value $x/2 = 0.25$.

As mentioned before, the transition kernel Q does not admit a density with respect to the Lebesgue measure. As a consequence, our approach is theoretically not well-adapted to tackle the estimation of Q. Nevertheless, we apply our estimator in this unsuitable framework in order to analyze its asymptotic behavior. Despite this weakness, one may observe on Figure 3 that our method performs pretty well. The kernel function and the constants α and β have been chosen in the same way as in the previous example.

Acknowledgments. The author would like to thank his two PhD advisors, François Dufour and Anne Gégout-Petit, for their suggestions and helpful comments.

References

[1] Aven, T., and Jensen, U. *Stochastic models in reliability*, vol. 41 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, 1999.

[2] Azaïs, R., Dufour, F., and Gégout-Petit, A. Nonparametric estimation of the conditional density of the inter-jumping times for piecewise-deterministic Markov processes. *Preprint arXiv:1202.2212v2* (2012).

[3] Basu, A. K., and Sahoo, D. K. On Berry-Esseen theorem for nonparametric density estimation in Markov sequences. *Bull. Inform. Cybernet. 30*, 1 (1998), 25–39.

[4] Chafaï, D., Malrieu, F., and Paroux, K. On the long time behavior of the TCP window size process. *Stochastic Process. Appl. 120*, 8 (2010), 1518–1534.

[5] Chiquet, J., and Limnios, N. A method to compute the transition function of a piecewise deterministic Markov process with application to reliability. *Statist. Probab. Lett. 78*, 12 (2008), 1397–1403.

[6] Clémençon, S. J. M. Adaptive estimation of the transition density of a regular Markov chain. *Math. Methods Statist. 9*, 4 (2000), 323–357.

[7] Davis, M. H. A. *Markov models and optimization*, vol. 49 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London, 1993.

[8] Doukhan, P., and Ghindès, M. Estimation de la transition de probabilité d'une chaîne de Markov Doëblin-récurrente. Étude du cas du processus autorégressif général d'ordre 1. *Stochastic Process. Appl. 15*, 3 (1983), 271–293.

[9] Duflo, M. *Random iterative models*. Applications of Mathematics. Springer-Verlag, Berlin, 1997.

[10] Genadot, A., and Thieullen, M. Averaging for a fully coupled Piecewise Deterministic Markov Process in Infinite Dimensions. *Advances in Applied Probability 44*, 3 (2012).

[11] Hernández-Lerma, O., Esparza, S. O., and Duran, B. S. Recursive nonparametric estimation of nonstationary Markov processes. *Bol. Soc. Mat. Mexicana (2) 33*, 2 (1988), 57–69.

[12] Hernández-Lerma, O., and Lasserre, J. B. *Markov chains and invariant probabilities*, vol. 211 of *Progress in Mathematics*. Birkhäuser Verlag, Basel, 2003.

[13] Hu, J., Wu, W., and Sastry, S. *Modeling subtilin production in* bacillus subtilis *using stochastic hybrid systems*. In R. Alur and G.J. Pappas, editors, *Hybrid Systems: Computation and Control*, number 2993 in LNCS, Springer-Verlag, Berlin, 2004.

[14] Lacour, C. Adaptive estimation of the transition density of a Markov chain. *Ann. Inst. H. Poincaré Probab. Statist. 43*, 5 (2007), 571–597.

[15] Lacour, C. Nonparametric estimation of the stationary density and the transition density of a Markov chain. *Stochastic Process. Appl. 118*, 2 (2008), 232–260.

[16] Liebscher, E. Density estimation for Markov chains. *Statistics 23*, 1 (1992), 27–48.

[17] Masry, E., and Györfi, L. Strong consistency and rates for recursive probability density estimators of stationary processes. *J. Multivariate Anal. 22*, 1 (1987), 79–93.

[18] Meyn, S., and Tweedie, R. L. *Markov chains and stochastic stability*, second ed. Cambridge University Press, Cambridge, 2009.

[19] Rosenblatt, M. Density estimates and Markov sequences. In *Nonparametric Techniques in Statistical Inference (Proc. Sympos., Indiana Univ., Bloomington, Ind., 1969)*. Cambridge Univ. Press, London, 1970, pp. 199–213.

[20] Roussas, G. G. Nonparametric estimation in Markov processes. *Ann. Inst. Statist. Math. 21* (1969), 73–87.

[21] Van Ryzin, J. On strong consistency of density estimates. *Ann. Math. Statist. 40* (1969), 1765–1772.

[22] Yakowitz, S. Nonparametric density and regression estimation for Markov sequences without mixing assumptions. *J. Multivariate Anal. 30*, 1 (1989), 124–136.

E-mail address: `romain.azais@inria.fr`

INRIA Bordeaux Sud-Ouest, team CQFD, France and Université Bordeaux, IMB, CNRS UMR 5251, 200, Avenue de la Vieille Tour, 33405 Talence cedex, France.